

## A NUMERICAL METHOD FOR SINGULAR TWO POINT BOUNDARY VALUE PROBLEMS\*

*Dedicated to R. D. Richtmyer in Honor of his 65th Birthday*

D. C. BRABSTON† AND H. B. KELLER‡

**Abstract.** The numerical solution of boundary value problems for linear systems of first order equations with a regular singular point at one endpoint is considered. The standard procedure of expanding about the singularity to get a nonsingular problem over a reduced interval is justified in some detail. Quite general boundary conditions are included which permit unbounded solutions. Error estimates are given and some numerical calculations are presented to check the theory.

**1. Introduction.** We consider the numerical solution of linear two point boundary value problems with a regular singularity at one endpoint. In particular the class of problems is formulated as:

$$(1.1a) \quad \mathcal{L}\mathbf{y}(t) \equiv \mathbf{y}'(t) - A(t)\mathbf{y}(t) = \mathbf{f}(t), \quad 0 < t \leq 1;$$

$$(1.1b) \quad \mathcal{B}\mathbf{y}(t) \equiv \lim_{t \downarrow 0} [B_0(t)\mathbf{y}(t) + B_1\mathbf{y}(1) - \mathbf{b}(t)] = \mathbf{0};$$

$$(1.1c) \quad A(t) \equiv t^{-1}R + A_0(t).$$

Here  $\mathbf{y}(t)$ ,  $\mathbf{f}(t)$ ,  $\mathbf{b}(t)$  are  $n$ -vectors while  $R$ ,  $A_0(t)$ ,  $B_0(t)$ ,  $B_1$  are  $n \times n$  matrices. We assume  $A_0(t)$  is, say, analytic on  $(0, \delta_0]$  and sufficiently smooth on  $(0, 1]$ . Similarly  $B_0(t)$ ,  $\mathbf{f}(t)$  and  $\mathbf{b}(t)$  are smooth on  $(0, 1]$  but may be singular at  $t = 0$ . The solution  $\mathbf{y}(t) \in C^1(0, 1]$  but need not even be bounded as  $t \rightarrow 0$ . The form of the boundary conditions (1.1b) includes the typical constraints of wave propagation problems (i.e. incoming or outgoing waves) and those due to symmetry in singular coordinate systems. With  $B_0(x) = \tilde{B}_0 t^{-R}$  conditions can be imposed on the regular part of the solution, as is done by Natterer [10]. That seems to be the first work to study numerical methods for fairly general linear systems of the form (1.1a, c).

We examine the more or less standard procedure of expanding about the singular point, solving a regular boundary value problem over a reduced interval excluding the singular point and matching this solution to the expansion. To study this process we require an existence and uniqueness theory for (1.1) which is developed in § 2. Then in § 3 we introduce a regular boundary value problem on some interval  $[\delta, 1]$  and show that it is equivalent to (1.1). The regular problem cannot be determined explicitly so in § 4 we study its replacement by a truncated regular problem which can be obtained explicitly. In § 5 we examine the error between the exact solution of (1.1) and a numerical solution of the truncated regular problem. Finally numerical examples are presented in § 6.

The extension of our methods to problems with two regular singular endpoints or even interior regular singularities is easily carried out, as shown in Brabston [1]. The study of unbounded intervals can be included if the point at

---

\* Received by the editors March 23, 1976, and in revised form July 23, 1976. This work was supported in part by ERDA under Grant AT-04-3-767.

† Data Systems Department, TRW Systems Group, Redondo Beach, California 90278.

‡ Applied Mathematics, California Institute of Technology, Pasadena, California 91125.

infinity is a regular singular point. This expansion technique or procedures very close to it have been used for many years. However, a study of the scalar case has only recently been made by Gustaffson [3]. Other studies of difference methods for special singular second order problems have been carried out by Jamet [5], Natterer [11], and Russell and Shampine [12]. A study of the trapezoidal and centered Euler difference schemes applied to bounded solutions of (1.1) and more general equations is in progress by de Hoog and Weiss [4].

**2. Existence and uniqueness theory.** A fundamental solution matrix,  $Y(t)$ , for (1.1a) can be defined, using some nonsingular  $n \times n$  matrix  $Y_0$ , by:

$$(2.1a) \quad \mathcal{L}Y(t) = 0, \quad 0 < t \leq 1;$$

$$(2.1b) \quad Y(\delta_0) = Y_0.$$

It is well known (see Coddington and Levinson [2, pp. 118–122]) that when  $R$  in (1.1c) has no eigenvalues separated by a positive integer:

$$(2.2a) \quad Y(t) = P(t)t^R, \quad 0 < t \leq \delta_0;$$

$$(2.2b) \quad P(t) = \sum_{k=0}^{\infty} P_k t^k, \quad P_0 = I, \dots;$$

where  $P(t)$  is analytic and nonsingular on  $0 \leq t \leq \delta_0$ . A simple transformation of (1.1a, c) insures the eigenvalue condition so we merely assume it to hold in all that follows (see, however, the example in § 2.1). Note that  $Y_0$  cannot be chosen arbitrarily in (2.1b) if we use the representation (2.2).

Every solution of (1.1a) can be written as

$$(2.3a) \quad \mathbf{y}(t) = Y(t)\mathbf{c} + \mathbf{y}_p(t), \quad 0 < t \leq 1;$$

where the particular solution,  $\mathbf{y}_p(t)$ , satisfies:

$$(2.3b) \quad \mathcal{L}\mathbf{y}_p(t) = \mathbf{f}(t), \quad 0 < t \leq 1, \quad \mathbf{y}_p(\delta_0) = \mathbf{y}_0.$$

In fact we have that

$$(2.3c) \quad \mathbf{y}_p(t) = Y(t) \left\{ \int_{\delta_0}^t Y^{-1}(\tau) \mathbf{f}(\tau) d\tau + Y^{-1}(\delta_0) \mathbf{y}_0 \right\}.$$

Thus  $\mathbf{y}(t)$ , given by (2.3a), will satisfy the boundary conditions (1.1b) if and only if:

$$(2.4a) \quad \lim_{t \downarrow 0} \{ [B(t) + B_1 Y(1)] \mathbf{c} - [\mathbf{g}(t) - B_1 \mathbf{y}_p(1)] \} = \mathbf{0}.$$

Here:

$$(2.4b) \quad B(t) \equiv B_0(t) Y(t);$$

$$(2.4c) \quad \mathbf{g}(t) \equiv \mathbf{b}(t) - B_0(t) \mathbf{y}_p(t).$$

So our existence and uniqueness theory is reduced to a study of the existence and uniqueness of a solution  $\mathbf{c}$ , of (2.4a).

Suppose, for the moment, that the possible singularities of  $B_0(t)$ ,  $\mathbf{b}(t)$  and  $\mathbf{f}(t)$  are of the forms:

$$(2.5) \quad B_0(t) \equiv \hat{B}_0(t)t^{R_0}, \quad \mathbf{b}(t) \equiv t^{R_1}\hat{\mathbf{b}}(t), \quad \mathbf{f}(t) \equiv t^{R_2}\hat{\mathbf{f}}(t),$$

where  $\hat{B}_0(t)$ ,  $\hat{\mathbf{b}}(t)$  and  $\hat{\mathbf{f}}(t)$  are analytic on  $[0, \delta_0]$ . Then using the Jordan forms of the  $n \times n$  matrices  $R, R_0, R_1$  and  $R_2$  with (2.2) and (2.5) in (2.4b, c) it easily follows that  $B(t)$  and  $\mathbf{g}(t)$  have a finite number of distinct singularities. More precisely we find, for some positive integers  $q$  and  $K$  that there exists a set of  $q$  scalar functions  $\varphi_\nu(t)$ ,  $\nu = 1, 2, \dots, q$ , satisfying:

$$(2.6a) \quad \lim_{t \downarrow 0} |\varphi_\nu(t)| = \infty,$$

$$(2.6b) \quad \lim_{t \downarrow 0} |\varphi_\nu(t)/\varphi_{\nu+1}(t)| = 0,$$

$$(2.6c) \quad \lim_{t \downarrow 0} t^K \varphi_q(t) = 0;$$

such that:

$$(2.6d) \quad B(t) = M_0(t) + \sum_{\nu=1}^q \varphi_\nu(t)M_\nu;$$

$$(2.6e) \quad \mathbf{g}(t) = \mathbf{g}_0(t) + \sum_{\nu=1}^q \varphi_\nu(t)\mathbf{g}_\nu.$$

Here  $M_0(t)$  and  $\mathbf{g}_0(t)$  are analytic on  $[0, \delta_0]$  while the  $M_\nu$  and  $\mathbf{g}_\nu$  are constant matrices and vectors. If  $B_0(t)$ ,  $\mathbf{b}(t)$  and  $\mathbf{f}(t)$  do not have the forms in (2.5) we require that their singularities be such that expansions of the form (2.6) are valid for some set  $\{\varphi_\nu(t)\}_1^q$ . Obviously finite sums of terms of the form used in (2.5) are easily included.

Now we have the basic existence and uniqueness result.

**THEOREM 2.7.** *Let the expansions (2.6) hold. Define the  $n(q+1)$  order matrix  $M$  and vector  $\mathbf{g}$  by:*

$$(2.7a) \quad M \equiv \begin{pmatrix} M_0(0) + B_1 Y(1) \\ M_1 \\ \vdots \\ M_q \end{pmatrix};$$

$$(2.7b) \quad \mathbf{g} \equiv \begin{pmatrix} \mathbf{g}_0(0) - B_1 \mathbf{y}_p(1) \\ \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_q \end{pmatrix}.$$

*Then (1.1a, b, c) has a solution if and only if  $\text{rank } M = \text{rank } (M, \mathbf{g})$ . The solution is unique if and only if  $\text{rank } M = n$ .*

*Proof.* We have shown above that our result is equivalent to the solvability of (2.4a). Using (2.6a, b, d, e) in (2.4a) we find that the coefficients of each singularity function,  $\varphi_\nu(t)$ , must vanish. This yields, in brief, the system

$$(2.7c) \quad M\mathbf{c} = \mathbf{g}.$$

Now the theorem follows from the elementary rank condition for linear systems.  $\square$

**2.1. An example.** It is easy to determine the singularities,  $\varphi_\nu(t)$ , and the constant quantities  $M_\nu$  and  $\mathbf{g}_\nu$ ,  $1 \leq \nu \leq q$ , of (2.6) explicitly when the forms (2.5) hold. In general, however, we cannot evaluate  $Y(1)$  or  $\mathbf{y}_p(1)$  so that  $M$  and  $\mathbf{g}$  may not be fully evaluated. An interesting exception is afforded by the inhomogeneous axisymmetric potential equation:

$$(2.8a) \quad v'' + \frac{\sigma}{t}v' = (\sigma - 2)t^{-\sigma} \sin t - t^{(1-\sigma)} \cos t, \quad 0 < t \leq 1, \quad \sigma > 0.$$

This or related forms are used by several authors to study numerical methods for singular points [3], [4], [5], [11], [12]. We write the equivalent first order system:

$$(2.8b) \quad \mathbf{y}' - \begin{pmatrix} 0 & 1 \\ 0 & -\sigma/t \end{pmatrix} \mathbf{y} = \begin{pmatrix} 0 \\ (\sigma - 2)t^{-\sigma} \sin t - t^{1-\sigma} \cos t \end{pmatrix}, \quad \mathbf{y} \equiv \begin{pmatrix} v \\ v' \end{pmatrix},$$

and consider boundary conditions of the general form (1.1b). Since  $\sigma = 1$  yields eigenvalues of  $R$  which differ by an integer it is a special case. However we easily find that the fundamental solution matrix for (2.8b) is:

$$(2.9a) \quad Y(t) = \begin{pmatrix} 1 & s_\sigma(t) \\ 0 & t^{-\sigma} \end{pmatrix}, \quad \text{where } s_\sigma(t) \equiv \begin{cases} \ln t, & \sigma = 1; \\ t^{1-\sigma}(1-\sigma)^{-1}, & \sigma \neq 1. \end{cases}$$

A particular solution of (2.8b) is, for all  $\sigma$ :

$$(2.9b) \quad \mathbf{y}_p(t) = \begin{pmatrix} t^{1-\sigma} \cos t \\ (1-\sigma)t^{-\sigma} \cos t - t^{1-\sigma} \sin t \end{pmatrix}.$$

For constant  $B_0(t) \equiv (B_{ij}^0)$  and  $\mathbf{b}(t) \equiv (b_1, b_2)^T$  we use (2.9) in (2.4b, c) to get the  $\{\phi_\nu(t)\}_1^q$  of (2.6) as:

$$(2.9c) \quad \begin{aligned} 0 < \sigma < 1: q = 1, & \quad \phi_1(t) = t^{-\sigma}; \\ \sigma = 1: q = 2, & \quad \phi_1(t) = \ln t, \quad \phi_2(t) = t^{-1}; \\ \sigma > 1: q = [\sigma] + 1, & \quad \phi_1(t) = t^{[\sigma]-\sigma}, \dots, \phi_{q-1}(t) = t^{1-\sigma}, \quad \phi_q(t) = t^{-\sigma}. \end{aligned}$$

Here  $[\sigma]$  is the largest integer less than  $\sigma$ . With  $B_1 \equiv (B_{ij}^1)$  we easily obtain the matrices  $M_\nu$  and vectors  $\mathbf{g}_\nu$  that enter into (2.7). Rather than list these quantities we state some of the conclusions that they imply; more details can be found in Brabston [1].

A. For  $0 < \sigma < 1$  a unique solution exists for every constant  $\mathbf{b}(t) \equiv \mathbf{b}$  provided:

$$B_0 = \begin{pmatrix} B_{11}^0 & 0 \\ B_{21}^0 & 0 \end{pmatrix} \quad \text{and} \quad M_0(0) + B_1 Y(1) = \begin{pmatrix} B_{11}^0 + B_{11}^1 & B_{11}^1(1-\sigma)^{-1} + B_{12}^1 \\ B_{21}^0 + B_{21}^1 & B_{21}^1(1-\sigma)^{-1} + B_{22}^1 \end{pmatrix}$$

is nonsingular.

In terms of the scalar formulation (2.8a) this special form of  $B_0$  implies that  $v'(t)$  cannot enter into the boundary conditions at  $t = 0$ .

B. For  $\sigma \geq 1$  a unique solution exists for every constant  $\mathbf{b}(t) \equiv \mathbf{b}$  provided:  $B_0 \equiv 0$  and  $B_1$  is nonsingular, that is an initial value problem must be posed with initial point  $t = 1$ .

Of course many special solutions exist for  $B_0$  and  $B_1$  which violate the above general conditions. For example

$$B_0 \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B_1 \equiv \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ (\sigma - 1)(1 - \cos 1) - \sin 1 \end{pmatrix}$$

correspond to the boundary conditions

$$v(0) = 1, \quad \lim_{t \downarrow 0} v'(t) + v'(1) = (\sigma - 1)(1 - \cos 1) - \sin 1.$$

For  $0 < \sigma < 1$  this yields the unique solution

$$v(t) = 1 - t^{1-\sigma}(1 - \cos t).$$

**3. Reduction to a regular problem.** We show how to replace the singular problem (1.1) by a regular problem on a reduced interval  $[\delta, 1]$ . Suppose (1.1) has a solution under the hypothesis of Theorem 2.7. This solution must have the form (2.3a) for some fixed  $\mathbf{c}$ . Then for any point  $\delta \in (0, \delta_0]$  since  $Y(\delta)$  is nonsingular, the vector  $\mathbf{c}$  is given by:

$$(3.1) \quad \mathbf{c} = Y^{-1}(\delta)[\mathbf{y}(\delta) - \mathbf{y}_p(\delta)].$$

But this vector  $\mathbf{c}$  must satisfy (2.7c) and thus we get that

$$(3.2a) \quad B_{0\delta}\mathbf{y}(\delta) + B_{1\delta}\mathbf{y}(1) = \mathbf{b}_\delta$$

where we have used  $\mathbf{y}(1) = Y(1)\mathbf{c} + \mathbf{y}_p(1)$  and introduced:

$$(3.2b) \quad B_{0\delta} \equiv \begin{pmatrix} M_0(0) \\ M_1 \\ \vdots \\ M_q \end{pmatrix} Y^{-1}(\delta), \quad B_{1\delta} \equiv \begin{pmatrix} B_1(0) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{b}_\delta \equiv \begin{pmatrix} \mathbf{g}_0(0) \\ \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_q \end{pmatrix} + B_{0\delta}\mathbf{y}_p(\delta).$$

It follows that the regular two point problem consisting of

$$(3.3) \quad \mathcal{L}\mathbf{u}(t) \equiv \mathbf{u}(t) - A(t)\mathbf{u}(t) = \mathbf{f}(t), \quad \delta \leq t \leq 1$$

subject to (3.2) has a solution when (1.1) has a solution given by (2.3a). Indeed we have even more as follows.

**THEOREM 3.4.** *Let  $\mathbf{y}_p(t)$  be a particular solution of (1.1a), say as defined in (2.3b). Then the singular problem (1.1) has a unique solution if and only if the regular problem (3.2)–(3.3) has a unique solution. Further, the unique solutions of both problems are identical on  $[\delta, 1]$ .*

*Proof.* If either problem has a solution it can be represented by (2.3a) with possibly differing constant vectors  $\mathbf{c}$ . Using this form in (3.2a) yields the linear system

$$(3.5) \quad [B_{0\delta}Y(\delta) + B_{1\delta}Y(1)]\mathbf{c} = \mathbf{b}_\delta - [B_{0\delta}\mathbf{y}_p(\delta) + B_{1\delta}\mathbf{y}_p(1)].$$

However, this system is precisely (2.7c). Thus both problems lead to the identical linear system and our results follow.  $\square$

**4. The truncated regular problem.** To determine the matrix  $B_{0\delta}$  and inhomogeneous data  $\mathbf{b}_\delta$  of (3.2) for the reduced regular problem we require  $Y(\delta)$  and  $\mathbf{y}_p(\delta)$ . These are not generally known so we approximate them. Taking  $\delta$  sufficiently small we truncate the power series representation of  $P(t)$  in (2.2) to get, say:

$$(4.1a) \quad Y^N(t) \equiv P^N(t)t^R, \quad 0 < t \leq \delta;$$

$$(4.1b) \quad P^N(t) \equiv \sum_{k=0}^N P_k t^k.$$

Using  $Y^N(t)$  in place of  $Y(t)$  in (2.3c) we also define the truncated particular solution,

$$(4.1c) \quad \mathbf{y}_p^N(t) \equiv Y^N(t) \left\{ \int_{\delta_0}^t [Y^N(\tau)]^{-1} \mathbf{f}(\tau) d\tau + Y^N(\delta_0)^{-1} \mathbf{y}_0 \right\}.$$

Of course, as we shall see, any other approximations to  $Y(t)$  and  $\mathbf{y}_p(t)$  would do provided they are sufficiently accurate at  $t = \delta$ . In the indicated expansion technique we must take  $N$  sufficiently large so that all singularities are included; that is so that  $Y(t) - Y^N(t)$  and  $\mathbf{y}_p(t) - \mathbf{y}_p^N(t)$  are regular at  $t = 0$  and in fact vanish as  $t \downarrow 0$ . This can be assured, if the highest order singularity occurs in  $t^R$ , by taking  $N \geq \max \operatorname{Re} [-\lambda_k(R)] + 1$ . We shall in fact assume that

$$(4.2a) \quad \max_{0 \leq t \leq \delta} (\|Y(t) - Y^N(t)\|, \|\mathbf{y}_p(t) - \mathbf{y}_p^N(t)\|) \leq \Delta(N, \delta)$$

where  $\Delta(N, \delta) \downarrow 0$  as  $N \uparrow \infty$  for any fixed  $\delta \in (0, \delta_0]$ . We require  $N$  to be so large that

$$(4.2b) \quad \|Y^{-1}(\delta)\| \Delta(N, \delta) < 1.$$

Then  $Y^N(\delta)$  is nonsingular, by the Banach lemma, and

$$(4.2c) \quad \|Y^N(\delta)^{-1}\| \leq \frac{\|Y^{-1}(\delta)\|}{1 - \|Y^{-1}(\delta)\| \Delta(N, \delta)}.$$

We have already used this assumption in writing (4.1c).

In analogy with (2.4b, c) we define the truncated quantities:

$$(4.3a) \quad B^N(t) \equiv B_0(t) Y^N(t),$$

$$(4.3b) \quad \mathbf{g}^N(t) \equiv \mathbf{b}(t) - B_0(t) \mathbf{y}_p^N(t).$$

If  $B_0(t)$  and/or  $\mathbf{f}(t)$  are singular at  $t = 0$  we require that  $N$  is taken sufficiently large so that  $B(t) - B^N(t)$  and  $\mathbf{g}(t) - \mathbf{g}^N(t)$  vanish as  $t \downarrow 0$ . This is always possible when the assumptions (2.6) are valid. Indeed then  $N \geq K$  and the equality holds if  $B_0(t)$  and

$\mathbf{f}(t)$  are regular. Under the above assumptions it follows that

$$(4.3c) \quad B^N(t) = M_0^N(t) + \sum_{\nu=1}^q \varphi_\nu(t) M_\nu, \quad M_0^N(0) = M_0(0);$$

$$(4.3d) \quad \mathbf{g}^N(t) = \mathbf{g}_0^N(t) + \sum_{\nu=1}^q \varphi_\nu(t) \mathbf{g}_\nu, \quad \mathbf{g}_0^N(0) = \mathbf{g}_0(0).$$

Here  $M_0^N(t)$  and  $\mathbf{g}_0^N(t)$  are “truncations” of  $M_0(t)$  and  $\mathbf{g}_0(t)$ , respectively, in (2.6d, e). Note that the  $M_\nu$ ,  $\mathbf{g}_\nu$  and  $\varphi_\nu(t)$  are retained exactly while the value of  $M_0^N(t)$  and  $\mathbf{g}_0^N(t)$  at  $t=0$  are exactly the values of  $M_0(0)$  and  $\mathbf{g}_0(0)$ , respectively.

A truncated regular boundary value problem is now defined as:

$$(4.4a) \quad \mathcal{L}\mathbf{y}^N(t) = \mathbf{f}(t), \quad \delta \leq t \leq 1,$$

$$(4.4b) \quad B_{0\delta}^N \mathbf{y}^N(\delta) + B_{1\delta} \mathbf{y}^N(1) = \mathbf{b}_\delta^N.$$

Here we recall (3.2b) and that  $Y^N(\delta)$  is nonsingular to introduce:

$$(4.4c) \quad B_{0\delta}^N \equiv B_{0\delta} Y(\delta) Y^N(\delta)^{-1}; \quad \mathbf{b}_\delta^N \equiv \mathbf{b}_\delta + B_{0\delta} \mathbf{y}_p^N(\delta) - B_{0\delta} \mathbf{y}_p(\delta).$$

We note that the truncated regular problem (4.4) does not employ  $Y(\delta)$  or  $\mathbf{y}_p(\delta)$  in its specification of (4.4c) but only  $Y^N(\delta)$  and  $\mathbf{y}_p^N(\delta)$ . Unfortunately even for very large  $N$  the truncated and untruncated problems need not be equivalent when  $q \geq 1$ . Indeed since any solution of (4.4a) can be represented as

$$(4.5) \quad \mathbf{y}^N(t) = Y(t) \mathbf{c}^N + \mathbf{y}_p(t), \quad \delta \leq t \leq 1,$$

it will be a solution of (4.4b) provided the constant  $\mathbf{c}^N$  satisfies:

$$(4.6a) \quad M^N \mathbf{c}^N = \mathbf{g}^N$$

where, if we recall (2.7a, b), (3.2b) and (4.4c):

$$(4.6b) \quad \begin{aligned} M^N &= B_{0\delta}^N Y(\delta) + B_{1\delta} Y(1) \\ &= M Y^N(\delta)^{-1} Y(\delta) + B_{1\delta} Y(1) [I - Y^N(\delta)^{-1} Y(\delta)] \end{aligned}$$

$$(4.6c) \quad \mathbf{g}^N = \mathbf{g} + B_{0\delta} [\mathbf{y}_p^N(\delta) - \mathbf{y}_p(\delta)].$$

Thus (4.4) has a solution if and only if  $\text{rank } M^N = \text{rank } (M^N, \mathbf{g}^N)$  and a solution is unique if  $\text{rank } M^N = n$ . We recall that (3.2)–(3.3) has a unique solution if and only if (2.7c) has a unique solution, that is  $\text{rank } M = \text{rank } (M, \mathbf{g}) = n$ . If  $\|Y^N(\delta) - Y(\delta)\|$  is small but nonzero while  $\mathbf{y}_p^N(\delta) = \mathbf{y}_p(\delta)$  the linear systems (4.6a) and (2.7c) are still not in general equivalent. They would be in this case if  $B_{1\delta} Y(1) [I - Y^N(\delta)^{-1} Y(\delta)] \equiv 0$ .

However when (3.2)–(3.3) has a unique solution we can be assured that a subset of the boundary conditions in (4.4b) does yield, for  $N$  sufficiently large, a uniquely solvable problem for (4.4a). More precisely we state this as follows.

**THEOREM 4.7.** *Let (1.1) have a unique solution and the expansions (2.6) hold. Then there exists some  $n \times n(q+1)$  projection matrix  $P$  such that  $PM$  is nonsingular. For any such  $P$  the problem:*

$$(4.7a) \quad \mathcal{L}\mathbf{y}^N(t) = \mathbf{f}(t), \quad \delta \leq t \leq 1;$$

$$(4.7b) \quad P[B_{0\delta}^N \mathbf{y}^N(\delta) + B_{1\delta} \mathbf{y}^N(1) - \mathbf{b}_\delta^N] = \mathbf{0}$$

*has a unique solution for all  $N$  sufficiently large.*

*Proof.* Since (1.1) has a unique solution, then by Theorem 2.7 the matrix  $M$  has rank  $n$ . So some  $n \times n$  submatrix of  $M$ , say  $PM$ , must be nonsingular; this yields the existence of  $P$ .

A solution of (4.7a) must have the form (4.5). It easily follows using (4.6) that (4.7b) will be satisfied if and only if

$$(4.8a) \quad PM^N \mathbf{c}^N = P\mathbf{g}^N.$$

However, recalling (4.2), we have

$$(4.8b) \quad \|PM - PM^N\| \leq K_0 \frac{\|Y^{-1}(\delta)\|\Delta(N, \delta)}{1 - \|Y^{-1}(\delta)\|\Delta(N, \delta)}$$

where  $K_0 = \|(M_0^T(0), M_1^T, \dots, M_q^T)^T\|$ . Then  $PM^N$  is nonsingular for  $\Delta(N, \delta)$  sufficiently small.  $\square$

By applying  $P$  to (2.7c) we easily conclude, under the above hypothesis and using (4.8), (4.4c), (4.2) and (3.2b), that for  $\Delta(N, \delta)$  sufficiently small:

$$(4.9a) \quad \|\mathbf{c}^N - \mathbf{c}\| \leq K_1(N, \delta)\Delta(N, \delta)$$

where

$$(4.9b) \quad K_1(N, \delta) \equiv \frac{\|(PM)^{-1}\|(1 + \|(PM)^{-1}\| \cdot \|P\mathbf{g}\|)K_0\|Y^{-1}(\delta)\|}{1 - [K_0\|(PM^{-1})\| + 1]\|Y^{-1}(\delta)\|\Delta(N, \delta)}.$$

We recall from (4.5), (2.3a) and Theorem 3.4 that  $\mathbf{y}(t)$  the unique solution of (1.1) and  $\mathbf{y}^N(t)$  the unique solution of the truncated problem (4.7) satisfy

$$(4.10) \quad \mathbf{y}^N(t) - \mathbf{y}(t) = Y(t)[\mathbf{c}^N - \mathbf{c}], \quad \delta \leq t \leq 1.$$

In actual applications we do not know  $M$  but only  $M^N$ . Thus we pick some  $P$  to assure  $PM^N$  is nonsingular for all  $N \geq N_0$ , say. In our experience it has always been  $Q = M_0(0) + B_1 Y(1)$ , the first  $n \times n$  submatrix of  $M$ , that is nonsingular.

**5. The numerical solution and error estimates.** When the singular problem (1.1) has a unique solution and (2.6) holds then we can determine a truncated regular two point problem (4.7) which also has a unique solution. There are many ways in which the solution,  $\mathbf{y}^N(t)$ , of this regular problem can be approximated. In particular we assume a stable accurate of order  $r$  difference scheme is used on a quasi-uniform net  $\{t_j\}_0^J$  with  $h = \max_j h_j$ :

$$(5.1) \quad t_0 = \delta; \quad t_j = t_{j-1} + h_j, \quad 1 \leq j \leq J; \quad t_J = 1.$$

(With fine net spacing near  $t_0$  we may also introduce netpoints in  $0 < t < \delta$  and on  $t > 1$  to employ deferred corrections in an efficient manner, if high order accuracy is required; see [8].) Denoting the numerical solution of (4.7) on the net (5.1) by  $\mathbf{u}_j$ , we have that:

$$(5.2) \quad \|\mathbf{y}^N(t_j) - \mathbf{u}_j\| \leq K_2(N, \delta)h^r, \quad 0 \leq j \leq J.$$

The theory which insures (5.2) is thoroughly developed; see for instance [9] for quite general difference schemes or [6] for the Box scheme employed in our calculations.



From (4.9), (4.10) and (5.2) we have that

$$(5.3) \quad \|\mathbf{y}(t_j) - \mathbf{u}_j\| \leq \|Y(t_j)\| K_1(N, \delta) \Delta(N, \delta) + K_2(N, \delta) h^r, \quad 0 \leq j \leq J, \quad \delta \leq t_j \leq 1.$$

The constant  $K_1(N, \delta)$  is given in (4.9b) while  $K_2(N, \delta)$  from (5.2) depends upon the magnitude of higher derivatives of  $\mathbf{y}^N(t)$  as well as the stability constant of the difference scheme over  $\delta \leq t \leq 1$  (see [6], [9] for more details). For *fixed*  $\delta$  both  $K_1(N, \delta)$  and  $K_2(N, \delta)$  are uniformly bounded for all  $N \geq N_0(\delta)$ , some sufficiently large integer. In most singular problems we can estimate  $\Delta(N, \delta)$  by

$$(5.4) \quad \Delta(N, \delta) = O(\delta^{N-q})$$

for some integer  $q$ . This can easily be done when the only singularities are those in  $Y(t)$ , see Brabston [1]. When (5.4) holds we need only take  $N - q = r$  to be compatible with any fixed difference scheme of accuracy  $h^r$  when  $h$  and  $\delta$  are comparable. However, for  $\delta$  too small  $K_1(N, \delta)$  of (4.9b) may degrade the accuracy if  $\|Y^{-1}(\delta)\| \Delta(N, \delta)$  is too large. Also for  $\delta = O(h)$  the stability constants of the difference scheme may cause  $K_2(N, \delta)$  to become unbounded. So this is not always a sure way to estimate the least value to use for  $N$ . But in general it is reasonable to take  $N > q + r$  in actual calculations. We see some of these effects in our calculations of § 6.

An approximation to the solution  $\mathbf{y}(t)$  of (1.1) on  $0 < t < \delta$  can be obtained from the finite difference solution by defining:

$$(5.5a) \quad \mathbf{c}_h^N \equiv [Y^N(\delta)]^{-1}(\mathbf{u}_0 - \mathbf{y}_p^N(\delta)),$$

$$(5.5b) \quad \mathbf{u}(t) \equiv Y^N(t) \mathbf{c}_h^N + \mathbf{y}_p^N(t), \quad 0 < t \leq \delta.$$

Here  $\mathbf{c}_h^N$  is defined in analogy with  $\mathbf{c}$  of (3.1) from the exact solution (2.3a). Thus we get, recalling (4.2) and (5.2) with  $t_0 = \delta$ :

$$(5.6a) \quad \|\mathbf{y}(t) - \mathbf{u}(t)\| \equiv \|\mathbf{e}(t)\| \leq \|Y(t)(\mathbf{c} - \mathbf{c}_h^N)\| + K_3(N, \delta) \Delta(N, \delta),$$

$$(5.6b) \quad \|\mathbf{c} - \mathbf{c}_h^N\| \leq K_4(N, \delta) \Delta(N, \delta) + K_5(N, \delta) h^r; \quad 0 < t \leq \delta.$$

If any row of  $Y(t)$  remains bounded as  $t \rightarrow 0$  the corresponding component of the error can be made arbitrarily small at  $t = 0$ . Otherwise our error bounds blow up as  $t \rightarrow 0$ . This should not be unexpected, however, as in our formulation we have allowed the exact solution to have singularities at  $t = 0$ . The relative error, say as measured by  $\|\mathbf{y}(t) - \mathbf{u}(t)\| / \|\mathbf{y}(t)\|$  as  $t \rightarrow 0$ , does remain bounded and can be made arbitrarily small on  $[0, \delta]$  as  $N \rightarrow \infty$  and  $h \rightarrow 0$ .

**6. Numerical examples.** We have computed approximations to the solutions of (2.8a) with  $\sigma$  values and boundary conditions given by:

$$(6.1a) \quad \sigma = \frac{1}{2}, \quad \lim_{t \downarrow 0} v(t) = 1, \quad v(1) = 2;$$

$$(6.1b) \quad \sigma = \frac{3}{2}, \quad \lim_{t \downarrow 0} v(t) = 1, \quad v(1) = \cos 1.$$

By our analysis in § 2 it follows that a unique solution exists in each case. These are easily found to be:

$$(6.2a) \quad v(t) = 1 + (1 - \cos 1)t^{1/2} + t^{1/2} \cos t,$$

$$(6.2b) \quad v(t) = 1 - t^{-1/2} + t^{-1/2} \cos t.$$

Note that the derivatives of the solutions in both cases are singular at  $t = 0$ .

With the use of the formulation (2.8b) our boundary conditions become:

$$\lim_{t \downarrow 0} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \mathbf{y}(t) + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \mathbf{y}(1) = \mathbf{b}$$

where:  $\mathbf{b}^T = (1, 2)$  for  $\sigma = \frac{1}{2}$  and  $\mathbf{b}^T = (1, \cos 1)$  for  $\sigma = \frac{3}{2}$ .

Trivial Taylor expansions in (2.9) yield  $\mathbf{y}_p^N(t)$  and  $\mathbf{Y}^N(t)$  and we find that

$$(6.3) \quad \Delta(N, \delta) = \begin{cases} \frac{1}{1-\sigma} \delta^{1-\sigma}, & N = 0, \\ \frac{2N+5-\sigma}{(2N+4)!} \delta^{2N+3-\sigma}, & N \geq 1. \end{cases}$$

For  $\sigma > 1$  we must take  $N \geq (\sigma - 3)/2$  and  $N \geq 1$  in order to include all the singularities as required to insure (4.2). Proceeding as in § 4 we obtain the truncated reduced boundary conditions as in (4.4b, c). For  $\sigma = \frac{1}{2}$  these become:

$$(6.4a) \quad B_{0\delta}^N \equiv \begin{pmatrix} 1 & -2\delta \\ 0 & 0 \end{pmatrix}, \quad B_{1\delta} \equiv \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

$$\mathbf{b}_\delta^N \equiv \begin{pmatrix} 1 - 2 \sum_{k=0}^N \frac{(-1)^k}{(2k+1)!} \delta^{2k+5/2} \\ 2 \end{pmatrix}$$

and for  $\sigma = \frac{3}{2}$  they yield, upon dropping a pair of redundant conditions:

$$(6.4b) \quad B_{0\delta}^N \equiv \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B_{1\delta} \equiv \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

$$\mathbf{b}_\delta^N \equiv \begin{pmatrix} 1 - \sum_{k=0}^N \frac{(-1)^k}{(2k+2)!} \delta^{2k+3/2} \\ \cos 1 \end{pmatrix}.$$

Note that the  $B_{0\delta}^N$  are independent of  $N$  and hence turn out to be *exact* in this example; the  $\mathbf{b}_\delta^N$  are not exact, however.

We use the Box scheme (or centered Euler) as presented in [6] to solve the system (2.8b) subject to (4.4b) using the quantities in (6.4). The net was chosen to be uniform with spacing  $h = (1 - \delta)/J$ . The difference equations were solved on an IBM 370/158 in double precision using a stable block elimination procedure (not the one implied in [6] but rather case i) of eq. (5.7) in [7]). It takes approximately 476 milliseconds to solve one problem with  $J = 80$  intervals and the computing time is linear in  $J$ . The results thus computed are  $O(h^2)$  accurate approximations. We also used one Richardson extrapolation to get  $O(h^4)$  accurate approximations.

A particularly important and sensitive measure of the accuracy is  $\|\mathbf{c} - \mathbf{c}_h^N\|$ . In Table 1 this error is tabulated for both problems, for a sequence of refined nets both for the Box scheme (I) and for one extrapolation (II). We used  $\delta = 0.1$  and  $N = 6$  in this series of calculations. The theoretical estimate (5.6b) is in part borne out by these results. In particular since here  $\Delta(N, \delta) \approx O(10^{-14})$  the dominant

TABLE 1  
The errors,  $\|\mathbf{c} - \mathbf{c}_h^N\|$ , for varying nets with  $\delta = 0.1$  and  $N = 6$

$J$	$\sigma = \frac{1}{2}$		$\sigma = \frac{3}{2}$	
	$I(h^2)$	$II(h^4)$	$I(h^2)$	$II(h^4)$
10	.813 (.2)	.218 (-3)	—	—
20	.220 (-2)	.173 (-4)	.929 (-4)	.785 (-6)
40	.562 (-3)	.117 (-5)	.238 (-4)	.511 (-7)
60	.251 (-3)	.235 (-6)	.106 (-4)	.102 (-7)
80	.141 (-3)	.748 (-7)	.599 (-5)	.323 (-8)

term should be  $O(h^r)$  for  $r = 2$  or 4. Taking ratios as indicated we obtain the results in Table 2. The theoretically correct limiting ratios are 4 and 16. The unextrapolated calculations are relatively closer to the theory. This is probably caused by the  $\Delta(N, \delta)$  term whose influence is greater for the more accurate  $O(h^4)$  results.

TABLE 2  
Ratios of errors  $\|\mathbf{c} - \mathbf{c}_h^N\|$  for successively bisected nets with  $J_1$  and  $J_2$  intervals each

$J_1/J_2$	$\sigma = \frac{1}{2}$		$\sigma = \frac{3}{2}$	
10/20	3.70	12.60	—	—
20/40	3.91	14.79	3.90	15.36
40/80	3.99	15.64	3.97	15.82

To assess this term we vary  $N$  while keeping  $h$  fixed with  $J = 80$  and  $\delta$  fixed at  $10^{-1}$  for the  $\sigma = \frac{1}{2}$  case. The results using one Richardson extrapolation are as follows:

$N$ :	0	1	2	3
$\ \mathbf{c} - \mathbf{c}_h^N\ $ :	.134 (-4)	.812 (-7)	.748 (-7)	.748 (-7)

If the  $O(h^r)$  term in (5.6b) is negligible, with  $r = 4$  since we use extrapolation these values should vary as  $\Delta(N, \delta)$ . From (6.3) we see that this does occur in going from  $N = 0$ ,  $\Delta(0, \delta) = O(\delta^{1/2})$  to  $N = 1$ ,  $\Delta(1, \delta) = O(\delta^{9/2})$ . After that the increase in  $N$ , with  $\Delta(N, \delta) = O(\delta^{(4N+5)/2})$  is apparently masked by the truncation error which is  $O(10^{-8})$  with  $J = 80$ .

The variation with  $\delta$  is more difficult to check since the variation of the coefficients,  $K_\nu(N, \delta)$ , in (5.6b) is not known and may be quite sensitive. Using the Box scheme, with  $O(h^2)$  accuracy,  $\sigma = \frac{1}{2}$ ,  $N = 6$  and  $J = 80$  we find the following

$\delta$ :	$5 \times 10^{-2}$	$10^{-1}$	$2 \times 10^{-1}$	$4 \times 10^{-1}$
$\ \mathbf{c} - \mathbf{c}_h^N\ $ :	.636 (-3)	.141 (-3)	.258 (-4)	.279 (-5)

If, as is likely, the  $\Delta(N, \delta)$  term is negligible compared to the  $O(h^2)$  term then ratios of successive terms should be proportional to  $[(1 - \delta_\nu)/(1 - \delta_{\nu+1})]^2$  since

$h_\nu = (1 - \delta_\nu)/80$  in each case. This assumes that  $K_4(N, \delta)$  does not vary much with  $\delta$ . The ratio of errors gives:

$$4.59, \quad 5.47, \quad 9.25$$

while the squares of the ratio of net spacings gives:

$$1.114, \quad 1.266, \quad 1.778.$$

Thus there is no satisfactory agreement here and we assume that the coefficient variations with  $\delta$  cannot be neglected.

TABLE 3  
The errors in  $v(t)$  and  $v'(t)$  over  $0 < t < \delta = 10^{-1}$

$t$	$\sigma = \frac{1}{2}$		$\sigma = \frac{3}{2}$	
	$ e_1(t) $	$ e_2(t) $	$ e_1(t) $	$ e_2(t) $
0.100	.894 (-4)	.447 (-3)	.139 (-15)	.300 (-4)
0.075	.775 (-4)	.515 (-3)	.915 (-6)	.459 (-4)
0.051	.635 (-4)	.629 (-3)	.244 (-5)	.835 (-4)
0.016	.356 (-4)	.112 (-2)	.906 (-5)	.475 (-3)
0.006	.218 (-4)	.183 (-2)	.186 (-4)	.206 (-2)
0.001	.894 (-5)	.447 (-2)	.539 (-4)	.300 (-1)

Finally we examine the error over  $0 < t \leq \delta$  when (5.5) is used. The results are shown for both problems with  $\delta = 10^{-1}$ ,  $N = 6$  and  $J = 80$  in Table 3. Here  $|e_1(t)|$  and  $|e_2(t)|$  are the absolute errors in approximating  $v(t)$  and  $v'(t)$ , respectively. From (2.9a) we see that the first row of  $Y(t)$  is nonsingular for  $\sigma = \frac{1}{2}$  but it is singular for  $\sigma = \frac{3}{2}$ . The second row is singular in both cases. Then the error estimate in (5.6a) indicates that  $|e_1(t)|$  should remain bounded as  $t \rightarrow 0$  in the  $\sigma = \frac{1}{2}$  case and may become unbounded for  $\sigma = \frac{3}{2}$ . The error  $|e_2(t)|$  may become unbounded in both cases. Specifically as  $t \downarrow 0$ ,  $|e_1(t)|$  decays like  $t^{1/2}$  and  $|e_2(t)|$  grows like  $t^{-1/2}$  for  $\sigma = \frac{1}{2}$ . For  $\sigma = \frac{3}{2}$ ,  $|e_1(t)|$  grows like  $t^{-1/2}$  (but with some specious values) and  $|e_2(t)|$  grows like  $t^{-3/2}$ . This is consistent with (5.6a) and (2.9a).

#### REFERENCES

- [1] D. BRABSTON, *Numerical solution of singular endpoint boundary value problems*, part II of Ph.D. thesis, Calif. Inst. of Tech., Pasadena, 1974.
- [2] E. CODDINGTON AND N. LEVINSON, *Ordinary Differential Equations*, McGraw-Hill, New York, 1955.
- [3] B. GUSTAFSSON, *A numerical method for solving singular boundary value problems*, Numer. Math., 21 (1973), pp. 328-344.
- [4] F. DE HOOG AND R. WEISS, *Difference methods for boundary value problems with a singularity of the first kind*, MRC Tech. Summary Rep. 1536, Math. Res. Center, Univ. of Wisc., Madison, 1975.
- [5] P. JAMET, *On the convergence of finite difference approximation to one-dimensional singular boundary-value problems*, Numer. Math., 14 (1970), pp. 355-378.
- [6] H. B. KELLER, *Accurate difference methods for linear ordinary differential systems subject to linear constraints*, this Journal, 6 (1969), pp. 8-30.

- [7] ———, *Accurate difference methods for nonlinear two-point boundary value problems*, this Journal, 11 (1974), pp. 305–320.
- [8] H. B. KELLER AND V. PEREYRA, *Deferred corrections for two-point boundary value problems*, in preparation.
- [9] H. B. KELLER AND A. WHITE, *Difference methods for boundary value problems in ordinary differential equations*, this Journal, 12 (1975), pp. 791–802.
- [10] F. NATTERER, *A generalized spline method for singular boundary value problems of ordinary differential equations*, Linear Algebra Appl., 7 (1973), pp. 189–216.
- [11] ———, *Das Differenzenverfahren für singuläre Rand-Eigenwertaufgaben gewöhnlicher Differentialgleichungen*, Numer. Math., 23 (1975), pp. 387–409.
- [12] R. D. RUSSELL AND L. F. SHAMPINE, *Numerical methods for singular boundary value problems*, this Journal, 12 (1975), pp. 13–36.